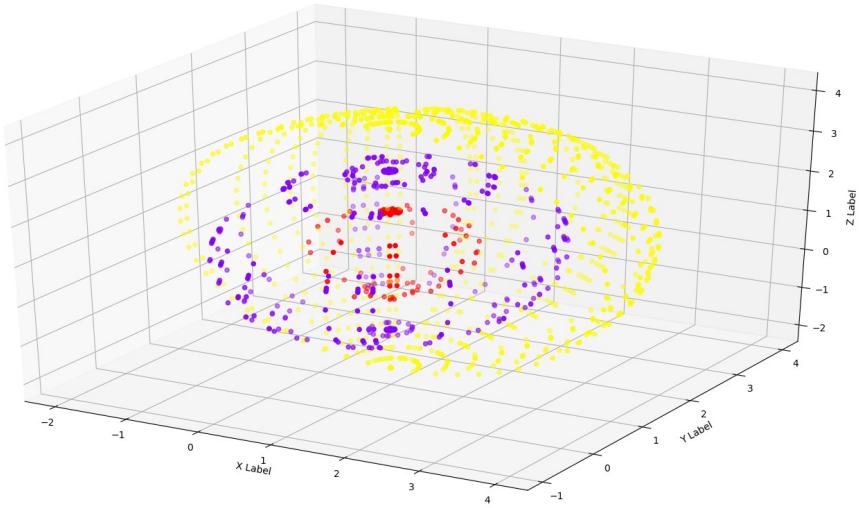
Machine Learning
Programming Assignment 2

Vaibhav Jindal
111701029

- Leaned how to use inbuilt functions in sklearn with different kernels.
- Difference between LDA and PCA practically.
- Got indepth knowledge of actual maths from the class and implementation using sklearn.
- Got a decent idea of 3-D plots in matplotlib.

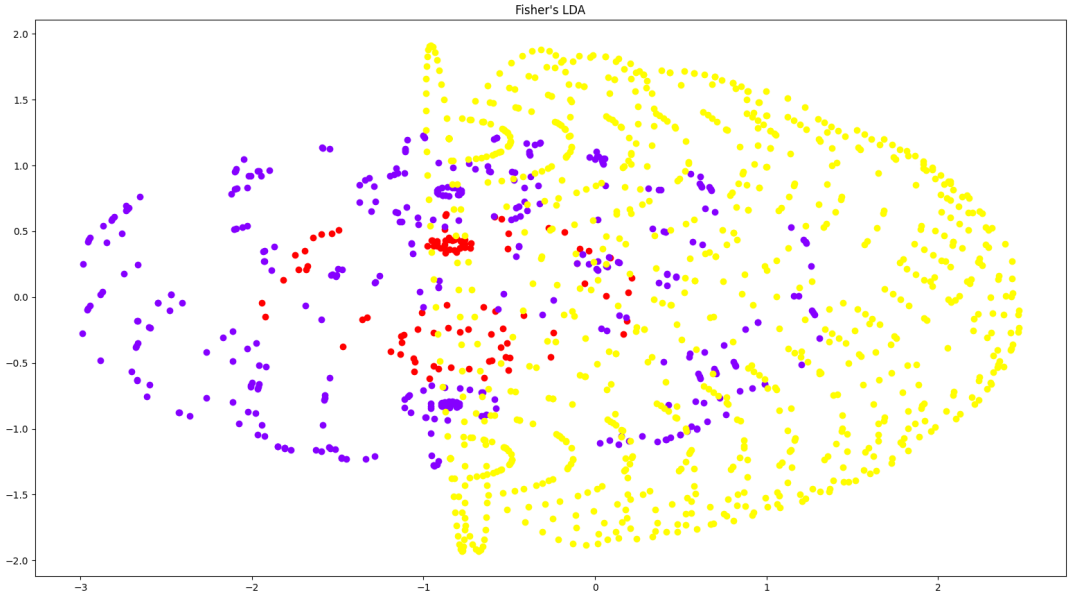
Taks 1: Generate data samples which are randomly distributed on 3 co-centric spheres in 3-D with centre at (1,1,1) and radius 1, 2 and 3 respectively. Plot a 3-D plot of them use 3 different colors to the points one 3 different spheres



Spheres

- Used parametric equation of a sphere in 3-D to make a sphere.
- Randomness is added through noise using np.random.normal (Noise is normal with mean 0 and variance 0.25)
- 100, 400 and 900 data_samples of each class is made respectively (smaller sphere to bigger sphere)
- scatter3D function used to print 3-D plot

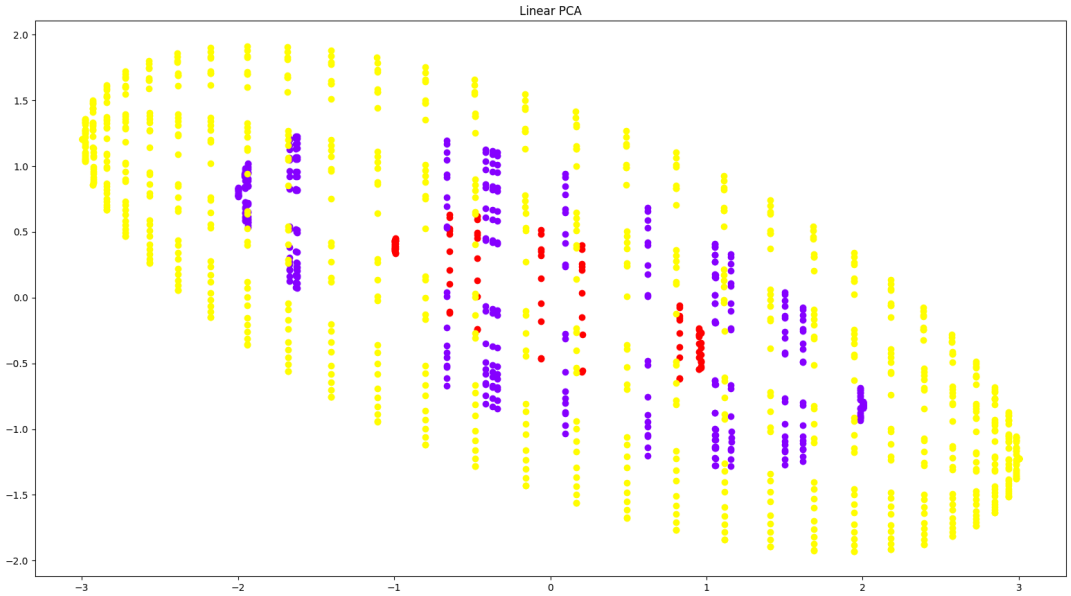
Taks 2: Find 2-D representation of these samples using Linear PCA, LDA, PCA with polynomial kernel with degree 5 and PCA with Gaussian kernel.



Fisher’s LDA

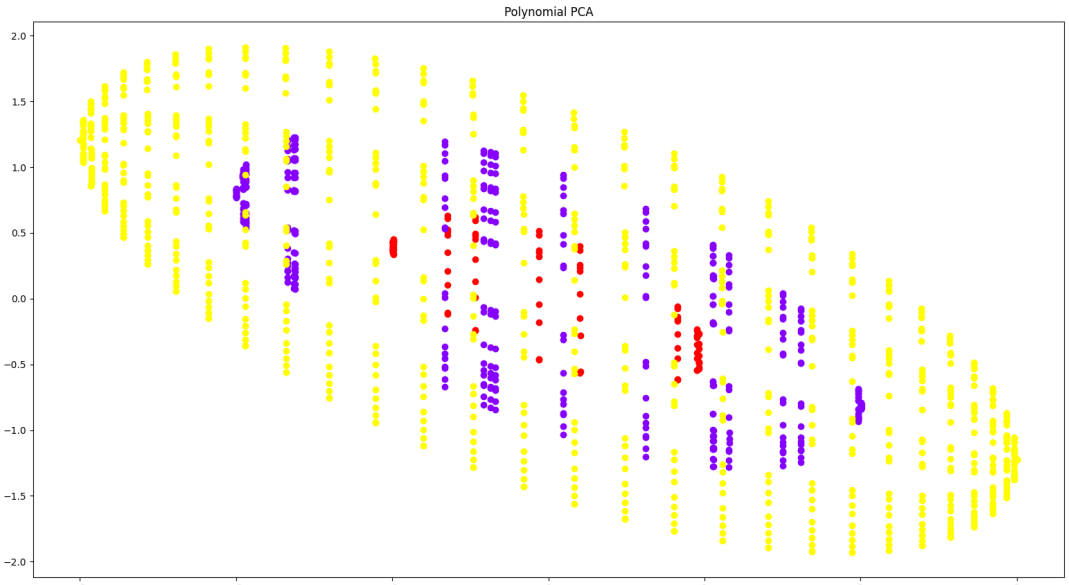
Linear Discriminant Analysis is a supervised algorithm as it takes the class label into consideration. It reduce ‘dimensionality’ while at the same time preserving as much of the class discrimination information as possible.

- It maximize the distance between the mean of each classes. (Here it is 0)
- It minimize the variation, within each class category.



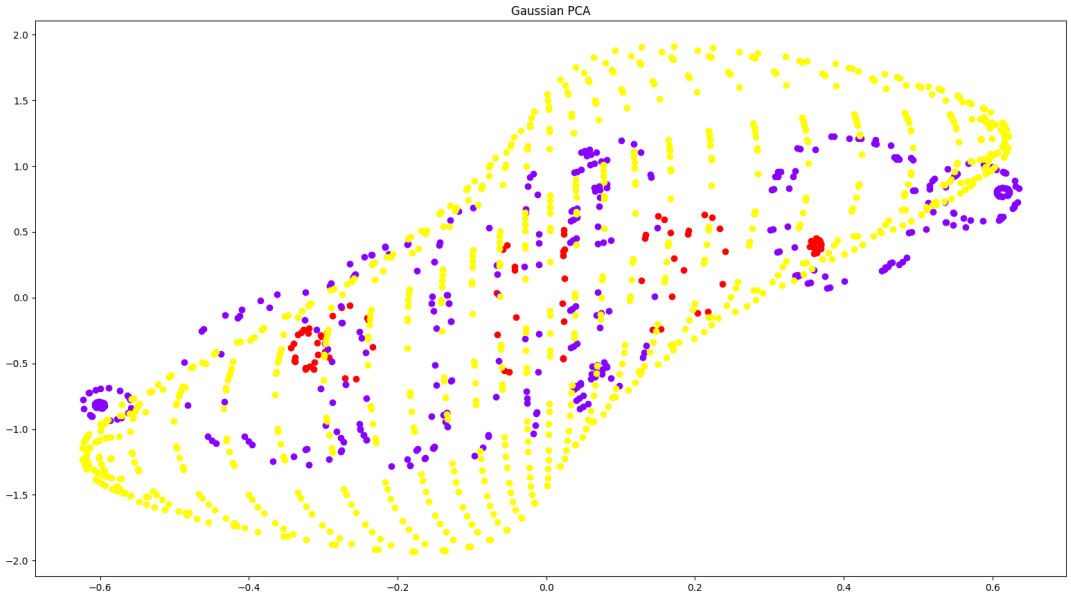
Linear PCA

- PCA is an unsupervised learning as it ignores the class labels.
- PCA gives the direction where the variance is maximum independent of the class.
- PCA determines the lines of variance in the dataset which are called as principal components with the first principal component having the maximum variance, second principal component having second maximum variance and so on.



Polynomial PCA

- Polynomial PCA doesn't have much difference from Linear PCA. Hence $(u.v + 1)^5$ ~= (u.v) here.



Gaussian PCA

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

- RBF kernel map the data into a infinite dimensional space.
- Gaussian kernel sees variance to more in the other direction ($+45^O$) different from linear and polynomial kernel.
- Gaussina kernel is trying to make individual circles in 2D.

CONCLUSION: PCA performs better in case where number of samples per class is less. Whereas LDA works better with large dataset having multiple classes; class separability is an important factor while reducing dimensionality.